Misleading funnel plot for detection of bias in meta-analysis

Jin-Ling Tang\textsuperscript{a,}* , Joseph LY Liu\textsuperscript{b}

\textsuperscript{a}Department of Community and Family Medicine, Faculty of Medicine, The Chinese University of Hong Kong, Shatin, N.T., Hong Kong
\textsuperscript{b}Centre for Clinical Trials and Epidemiological Research, Faculty of Medicine, The Chinese University of Hong Kong, Shatin, N.T., Hong Kong

Received 23 April 1999; received in revised form 24 August 1999; accepted 6 October 1999

Abstract
Publication and other forms of selection biases pose a threat to the validity of meta-analysis. Funnel plots are usually used to detect such biases; asymmetrical plots are interpreted to suggest that biases are present. Using 198 published meta-analyses, we demonstrate that the shape of a funnel plot is largely determined by the arbitrary choice of the method to construct the plot. When a different definition of precision and/or effect measure were used, the conclusion about the shape of the plot was altered in 37 (86\%) of the 43 meta-analyses with an asymmetrical plot suggesting selection bias. In the absence of a consensus on how the plot should be constructed, asymmetrical funnel plots should be interpreted cautiously. These findings also suggest that the discrepancies between large trials and corresponding meta-analyses and heterogeneity in meta-analyses may also be determined by how they are evaluated. © 2000 Elsevier Science Inc. All rights reserved.

Keywords: Funnel plot; Meta-analysis; Randomized controlled trials; Selection bias; Publication bias; Statistical method; Systematic reviews

1. Introduction
Meta-analysis is increasingly used to summarize health-related evidence as an aid to medical decision making. One of the main criticisms about the technique is that it is prone to publication and other forms of selection biases [1–5]. Meta-analyses should therefore be routinely scrutinized for the possible presence of such biases in practice [6]. The funnel plot (with the associated significance test for asymmetry of the plot) has been advocated by many for this purpose [1,6–8]. However, the procedure lacks empirical evidence to support what is observed is, indeed, selection bias. It was suggested that the method may result in false-positive results [9]. The validity of funnel plots needs to be scrutinized.

A funnel plot is constructed by plotting the trial-specific effect against a measure of its precision. Precision may be defined differently according to the number of subjects in the trial or a function of the standard error for the effect, such as the inverse of the standard error and the statistical power [1,6,10,11]. A commonly held interpretation is that the points will be symmetrically distributed around the overall effect of the meta-analysis; when selection bias is present, the plot will become asymmetrical and the overall effect of the meta-analysis will be biased [1,6]. However, this reasoning is overly simplistic because other factors may also affect the shape of the funnel plot. As is well recognized, the asymmetry of the funnel plot may be due to true heterogeneity. In other words, less precise trials may show a different effect from that of more precise trials because they differ in their design, participants’ characteristics, regimen of treatment, etc., and thus are not measuring the same underlying effect [7,12–15].

Furthermore, the funnel plot can be constructed in different ways. The treatment effect may be expressed as the relative risk or risk difference [16–18]. Precision may also be defined differently, as mentioned above [1,6,10,11]. Little is known about whether the choice of the definition of precision and effect measure in constructing the funnel plot may affect the shape of the plot. Using published meta-analyses, this article examines whether differently constructed funnel plots may result in discrepancies in conclusions about the symmetry of the plot and how often this discrepancy occurs.

2. Methods
2.1. Materials and general methods
Meta-analyses were identified by search of every systematic review electronically published in the second issue of the 1998 Cochrane Database of Systematic Reviews [19]. Meta-analyses having trial(s) using continuous variables as the outcome or trials having a proportion of outcome events of 0\% or 100\% in either of the two compared groups were
Fig. 1. Funnel plots of the relative risk of three meta-analyses according to the definition of precision.
excluded. We also excluded meta-analyses with five or fewer trials. The arbitrary cutoff used here is based on the minimum number of six trials adopted in Egger et al’s paper in which the funnel plot method and the associated significance test are described in more detail [6]. A total of 198 meta-analyses were included in this analysis.

The precision is defined in two ways: one according to the inverse of the standard error and the other the trial size. We will call the funnel plot using the former definition of precision the standard error- or SE-based funnel plot and the latter the trial size- or N-based funnel plot. The measure of effect is expressed as both the relative risk and risk difference. Heterogeneity was tested by the method suggested by DerSimonian and Laird [20]. All the analyses were conducted using the Statistical Analysis System (version 6.11).

2.2. Standard error-based funnel plot and related asymmetry test

The SE-based funnel plot is constructed by plotting the effect measure (e.g., the natural logarithm of the relative risk) against the inverse of its standard error. Its asymmetry is tested by a significance test using the linear regression method suggested by Egger and colleagues [6]. Briefly, in the regression the standardized effect (STE), defined as the effect divided by its standard error, is regressed against the precision of the effect, defined as the inverse of the standard error (regression equation: $\text{STE} = \alpha + \beta \times \text{precision}$). The intercept $\alpha$ provides a quantitative measure of the asymmetry and is of major interest. The more the intercept deviates from zero, the more pronounced the asymmetry. Negative values will indicate that less precise studies have a more pronounced effect than more precise studies and may thus suggest selection bias. If the P-value of the intercept is 0.10 or smaller, the asymmetry is considered to be statistically significant and accepted.

2.3. Study size-based funnel plot and related asymmetry test

In the N-based funnel plot the effect is plotted against the number of subjects or study size. Using the same principles as in the SE-based regression [6,21,22], we suggest here an N-based linear regression in which the product ($E$) of the effect and the square root of the average number of subjects in the two compared groups of the trial ($n$) is regressed against the latter (regression equation: $E = \alpha + \beta \times \sqrt{n}$). The interpretation of the intercept $\alpha$ and its P-value is the same as that of the SE-based regression outlined above.

2.4. Comparison of differently constructed funnel plots

Five meta-analyses from the Cochrane Database [23–27] are used to demonstrate how the funnel plot of the same set of studies may differ from each other when different definitions of precision and/or effect measures are used. In order to show how often the funnel plots may differ from each other, we divided the 198 meta-analyses into the following three
Fig. 2. SE-based or N-based funnel plots of three meta-analyses according to the measure of effect.
groups: 1) consistently symmetrical: the funnel plot is symmetrical regardless of the definition of precision and effect measure used, 2) consistently asymmetrical: the funnel plot is asymmetrical regardless of the definition of precision and effect measure used, and 3) inconsistently asymmetrical: the funnel plot is asymmetrical when only some, but not all, combinations of the two definitions of precision and two effect measures are used. Group 3 meta-analyses may have a symmetrical or asymmetrical funnel plot depending on how the plot is constructed and are considered to have discrepant funnel plots. Meta-analyses of groups 2 and 3 are further divided into two groups according to the sign of intercept: 1) the intercept is negative in at least one of the asymmetrical funnel plots differently constructed and 2) the intercept is positive in all the asymmetrical funnel plots. The former could suggest selection bias while the latter would not.

3. Results

3.1. Comparison of funnel plots constructed with the same effect measure

Fig. 1 shows that the funnel plot for the relative risk may differ from each other when different definitions of precision are used in constructing the plot. In meta-analysis A [23] the SE-based funnel plot (Fig. 1, A1) is largely symmetrical and the P-value for the significance test of the asymmetry (i.e., the regression intercept) is 0.32, suggesting that the effect is not associated with precision and small trials have a relative risk similar to that of the large trials. Conversely, the N-based plot (Fig. 1, A2) is asymmetrical (P = 0.09 for the regression intercept, and the same for the P-values below) and shows that small trials have a much smaller effect than that of the large trials.

Meta-analysis B [24] shows the opposite. The SE-based plot (Fig. 1, B1) and test for the intercept (P = 0.10) suggest that small trials produce more pronounced effects than the large trials, while the N-based plot (Fig. 1, B2) and test (P = 0.24) suggest a lack of strong evidence for such an association and that the overall effect of small trials is close to that of the large ones.

In meta-analysis C [25] both the SE-based and N-based plots look asymmetrical but in opposite directions. The SE-based plot (Fig. 1, C1) indicates that small trials produce more pronounced effect than large ones (P = 0.09), while the N-based plot (Fig. 1, C2) seems to indicate the opposite, although the test for asymmetry was not statistically significant (P = 0.17).
3.2. Comparison of funnel plots constructed with the same definition of precision

Fig. 2 shows that funnel plots based on the same definition of large trials may differ from each other when different effect measures are used. The SE-based funnel plot of meta-analysis D [24] is asymmetrical for both the risk difference (Fig. 2, D1, P = 0.03) and relative risk (Fig. 2, D2, P = 0.10) but in opposite directions. The N-based funnel plot of meta-analysis E [26] is symmetrical for the risk difference (Fig. 2, E1, P = 0.91) but asymmetrical for the relative risk (Fig. 2, E2, P = 0.02).

In all five examples above, there is obvious heterogeneity in either the risk difference or the relative risk or both (P \( \geq \) 0.10). However, funnel plots may still differ in meta-analyses in which there is insufficient evidence for the presence of heterogeneity. In meta-analysis F [27], for example, heterogeneity is neither observed for the risk difference (P = 0.78) nor for the relative risk (P = 0.88). The SE-based funnel plot is symmetrical for the risk difference (Fig. 2, F1, P = 0.90) but asymmetrical for the relative risk (Fig. 2, F2, P = 0.06).

3.3. Frequency of meta-analyses having discrepant funnel plots

Table 1 shows how often the funnel plot differs when different definitions of precision and/or effect measures are used. Of the 198 meta-analyses, 122 (61.6%) consistently showed a symmetrical funnel plot (group 1), 8 (4.0%) an asymmetrical plot regardless of how the plot was constructed (groups 2 and 3), and 68 (34.3%) had discrepant funnel plots when a different definition of precision and/or effect measure were used (groups 4 and 5). The asymmetry of funnel plots is dependent on how the plot is constructed in 37 (86.0%, 95% confidence interval: 76.9–96.0%) (group 4) of the 43 meta-analyses with an asymmetrical funnel plot, which suggests selection bias (groups 2 and 4). The number is 31 (93.5%, 95% confidence interval: 86.1–100.0%) (group 5) of the 33 meta-analyses with an asymmetrical plot, which is unlikely to be a result of selection bias (groups 3 and 5).

The proportion of discrepant meta-analyses suggestive of selection bias is 18.5% (12/65) in meta-analyses with 6 trials, 21.4% (18/84) in those with 7–9 trials, and 26.5% (13/49) in those with 10 trials or more. In these discrepant meta-analyses, the proportion of those that are method-dependent is 100%, 88.9%, and 69.2% respectively, for the three groups of meta-analyses with different number of trials (data are not tabulated). In addition, in 95 meta-analyses in which there is no sufficient evidence for heterogeneity (P > 0.10) in either the risk difference or the relative risk, 23 (24.2%) show an asymmetrical funnel plot (i.e., equivalent to groups 2 and 4 meta-analyses in Table 1), suggesting selection bias, and in 19 (82.6%) of them the asymmetry is method dependent (data are not tabulated).

4. Discussion

We found that the shape of the funnel plot depends to a large degree on how the precision is defined and what effect measure is used in constructing the plot. The funnel plot is asymmetrical and suggests selection bias in 43 (21.7%) of the 198 meta-analyses examined in this study. However, the plot became symmetrical in 86.0% of the 43 meta-analyses when a different definition of precision and/or effect measure were used. This proportion was slightly smaller in meta-analyses with no sufficient evidence for heterogeneity and in those with a relatively large number of trials. In the absence of a consensus on which definitions of precision and effect measures are the preferred methods, this large number of inconsistently asymmetrical funnel plots should be interpreted cautiously.

Selection bias is possibly a reasonable explanation in the 6 of the 43 meta-analyses that consistently showed an asymmetrical funnel plot suggestive of selection bias. However, explanations other than selection bias may also need to be sought in the rest of the 43 meta-analyses. First, whether selection bias is present or not should not depend on how the funnel plot is constructed. Second, there were also 33 asymmetrical meta-analyses that have a positive intercept and don’t suggest selection bias. This may imply that a similar number of meta-analyses having a negative intercept suggestive of selection bias may also arise due to reasons other than selection bias. One explanation suggested is true heterogeneity or more precisely an association of the effect with the precision of the trials [7,12–15]. True heterogeneity, again, does not explain why most of the asymmetrical funnel plots are method dependent. The reasons for such dependence have yet to be found. We believe that the dependence can partly be explained by the variation in the frequency of the outcome events (or simply the risk) in the compared groups of the trials.

An examination of how the standard error is determined may explain this claim. The standard error is a function of the sample size and risk. Precision as defined by the inverse of standard error is thus a function of the sample size and risk, while precision as defined by the trial size depends by definition on sample size only. When the risk varies in compared groups among the trials, the precision of the same trials will
be rated differently by the two definitions of precision. As a result, the two funnel plots may differ. Meta-analyses in Fig. 1 provide cases in point when the precision is defined differently. In addition, the standard error is a different function of sample size and risk for the risk difference and the logarithm of the relative risk [16,18,28]. Once again, when the risk varies in the trials, the precision of the same trials will be rated differently when different effect measures are used, resulting in discrepant SE-based funnel plots (Fig. 2, D and F).

Furthermore, an asymmetrical funnel plot indicates that the treatment effect varies according to a third factor: the precision. This phenomenon is known as effect modification (or interaction) in epidemiology and the precision is the effect modifier [29]. The presence of effect modification largely depends on the choice of the effect measure when the risk in the control group differs [13,30,31]. Typically, no effect modification when the risk difference is used as the effect measure may indicate an effect modification when the relative risk is used, and vice versa [31]. Thus, the shape of funnel plots using different effect measures may differ in a manner similar to the dependence of effect modification on the effect measure used, even when trial size is used as precision (Fig. 2, E and Table 2).

To illustrate, assume that trials in a meta-analysis have the same number of patients equally divided in the two compared groups, and the risk in the treated group is 4% lower than that in the control group, which varies between 5% and 20% in the trials. The risk difference will be constant in the trials, but the relative risk varies; those with a larger relative risk will be rated as less precise than those with a smaller relative risk. Consequently, the SE-based funnel plot for the relative risk will be asymmetrical, indicating that less precise trials have a greater relative risk—a feature of publication bias but not a result of selective publication, whereas the SE-based funnel plot for the risk difference and the N-based plots will not suggest publication bias.

These discussions also indicate that the N-based funnel plots in which the precision is not affected by the risk may be the preferable method for detection of bias. However, in practice the trial size may not be independent of, but often determined by, the level of the risk studied [13,14]. Hence, like the SE-based funnel plots, the N-based funnel plots may cause false-positive results. Furthermore, if publication bias exists, the selection procedure is likely to be determined by a function of the magnitude of the effect and its standard error (as reflected by the associated confidence interval or P-value), rather than directly by study size [3]. In this case the N-based funnel plot may fail to detect the presence of the bias.

How often would the shape of funnel plots depend on the way the plot is constructed? The answer is whenever an asymmetrical funnel plot occurs. The variation in effect due to either true heterogeneity or chance is the prerequisite for the occurrence of an asymmetrical plot. Whenever the effect varies, the risk varies either in the control or treated group or in both. Thus, the risk variation, which is an inevitable consequence of effect variation, will always play a role, if not the only role, in the occurrence of an asymmetrical plot. In other words, all asymmetrical funnel plots are dependent more or less on the way they are constructed.

These findings raise a question about the validity of the funnel plot as a tool for the detection of selection-related biases in meta-analyses of binary outcomes because of its inability of distinguishing selection bias from true heterogeneity. Recognition of the dependence of the shape of the funnel plot on the definition of precision and/or effect measure used has several implications. First, it is overly simplistic to conclude that a meta-analysis is biased on an observation of an asymmetrical funnel plot. Second, most observed discrepancies between large trials and corresponding meta-analyses [32–34] may be dependent on how large trials are defined and what effect measures are used in the comparison. Third, the presence of heterogeneity in a meta-analysis may also depend on the choice of effect measure used [28]. Fourth, the effect measure that has a symmetrical funnel plot should probably be used to summarize the trials if it is also clinically desirable. Alternatively, an asymmetrical funnel plot should be interpreted as suggestive of a special form of heterogeneity in which the effect is associated with the precision (i.e., the risk, trial size or both) of the trials. This association can be simply called the precision-related heterogeneity. The determinants of the association, one of which is selection bias, should be further investigated and such information would be of great value for clinical decision making [35–39].

### Acknowledgments

We thank Shelly LY Tse and Lan Kwok for assistance in hand searching the Cochrane Database and Aprille Sham in...
statistical graphics. This project was supported in part by a grant from the Chinese University of Hong Kong (6900511).

References
